Catalogue no. 82-003-X
ISSN 1209-1367

Health Reports

Linking 2006 Census and hospital data in Canada

by Michelle Rotermann, Claudia Sanmartin, Richard Trudeau and Hélène St-Jean

Release date: October 21, 2015

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at infostats@canada.ca

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-877-287-4369

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard table symbols

The following symbols are used in Statistics Canada publications:

. not available for any reference period
.. not available for a specific reference period
... not applicable
0 true zero or a value rounded to zero
0^s value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p} preliminary
^{r} revised
x suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E} use with caution
F too unreliable to be published
* significantly different from reference category ($p < 0.05$)

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2015

All rights reserved. Use of this publication is governed by the Statistics Canada Open Licence Agreement.

An HTML version **is also available.**

Cette publication est aussi disponible en français.

Linking 2006 Census and hospital data in Canada

by Michelle Rotermann, Claudia Sanmartin, Richard Trudeau and Hélène St-Jean

Abstract

Background: Record linkage is commonly used in health research to fill data gaps. This study summarizes the linkage of the 2006 Census of Population (excluding Quebec) to hospital data from the Discharge Abstract Database (DAD).

Data and methods: Hierarchical deterministic exact matching was employed to link 2006 Census and DAD (2006/2007, 2007/2008 and 2008/2009) data, based on linkage keys derived from three variables common to both files—date of birth, postal code and sex. The full census file (short-form; 23.4 million) was used for record linkage; the 20% file (long-form; 4.65 million) representing the study cohort was used for validation. Linked files were compared across jurisdictions, years and other selected covariates in terms of eligibility for linkage, keys linked, and linkage and coverage rates.

Results: Overall, 80% of linkage keys identified in the DAD were linked to the 2006 Census. The percentage of long-form census respondents linked to at least one hospital record ranged between 5% and 8% across jurisdictions; linkage rates were higher among known high users of hospital services: older age groups, lower-income individuals, and Aboriginal people. In general, the linked census file represents the majority of hospital events that occurred during the study period. Coverage rates (weighted/unweighted) varied by geography and age group, with lower weighted rates for the territories and some younger age groups.

Interpretation: With hierarchical deterministic exact matching, census data can be linked to multiple years of DAD data. Incorporation of updated postal codes from tax files reduced linkage rate attrition over time. Lower coverage rates for the territories and younger age groups suggest that these populations may be underrepresented in the linked files.

Keywords: Coverage, data collection, data linkage, databases, medical records, hospital records

Record linkage, the process of matching records across or within databases, is commonly used in health research to fill data gaps[1-7] and create a dataset with broad applications.[6-11] Most health-related linkages in Canada have relied on health insurance numbers (HINs) from provincial registries, which are unique to individuals.[2,3,11,12] However, HINs are not available in most databases (for instance, mortality, census, tax), and access to provincial registries is limited.

In the absence of a unique identifier and/or registry, an alternative approach— hierarchical deterministic exact matching—can be used to link health administrative databases and other data sources.[13] This involves matching different combinations of co-occurring person-level information.[14-18] Statistics Canada demonstrated the validity of this approach by linking census and hospitalization data in two provinces (Ontario and Manitoba).[19] The use of birth date, sex and postal code to link files yielded results similar to those produced using HINs from the provincial health insurance registries.[19]

This study presents the results of a hierarchical exact matching approach to link the 2006 Census of Population with hospital data for all provinces and territories (excluding Quebec) from the 2006/2007-to-2008/2009 Discharge Abstract Database (DAD). The purpose is to determine if the Census–DAD linkage performed similarly in different jurisdictions, and if linkage and coverage rates declined as time passed since the census. The linkage was approved by Statistics Canada's Policy Committee.[20] Use of the linked data is governed by the Directive on Record Linkage.[21]

Data sources

2006 Census of Population

The 2006 Census collected information using short- and long-form questionnaires. Most households (80%) received the short form, which contained eight basic questions, including the birth date, sex and marital status of all household members. The remaining households (20%) received the long form, which contained an additional 53 questions on topics such as education, ethnicity, mobility, income and employment.[22] In some regions, all households were asked to complete the long form: Nunavut, Northwest Territories (excluding Yellowknife), Yukon (excluding Whitehorse,) and other Indian Reserves and settlements.[22] The census represents 95% to 97% of the population in the provinces and 93% to 94% of the population in the territories.[23]

For the purposes of record linkage, the complete census file (23.4 million), which contains both short- and long-form records was used. The long-form records (4.65 million) constitute the study cohort used for validation.

To make inferences about the Canadian population based on information from the long-form questionnaire, data are often weighted.[24] Sampling weights account for the survey design and the under- or overrepresentation of people with certain characteristics.[25] The census weights were not adjusted for linkage eligibility.

Authors: Michelle Rotermann (Michelle.Rotermann@statcan.gc.ca) and Claudia Sanmartin are with the Health Analysis Division, Richard Trudeau (Richard.Trudeau@statcan.gc.ca) is with Special Surveys Division, and Hélène St-Jean is with the Household Survey Methods Division at Statistics Canada, Ottawa, Ontario.

Quality standards for census collection and processing were rigorous.[22] As well, census information underwent quality verification, including comparisons with alternative data sources. Inconsistent or missing responses were imputed to ensure the internal consistency of information provided by each household. The overall imputation rate was 2.9%; imputation rates for "age" and "sex" were less than 1.5%.

The accuracy of address information, including postal codes, was central to the success of data collection. In areas where census forms were delivered by mail, Statistics Canada validated and updated address information before Census Day (May 16, 2006). In areas where enumerators delivered the forms, addresses were listed and verified at delivery.

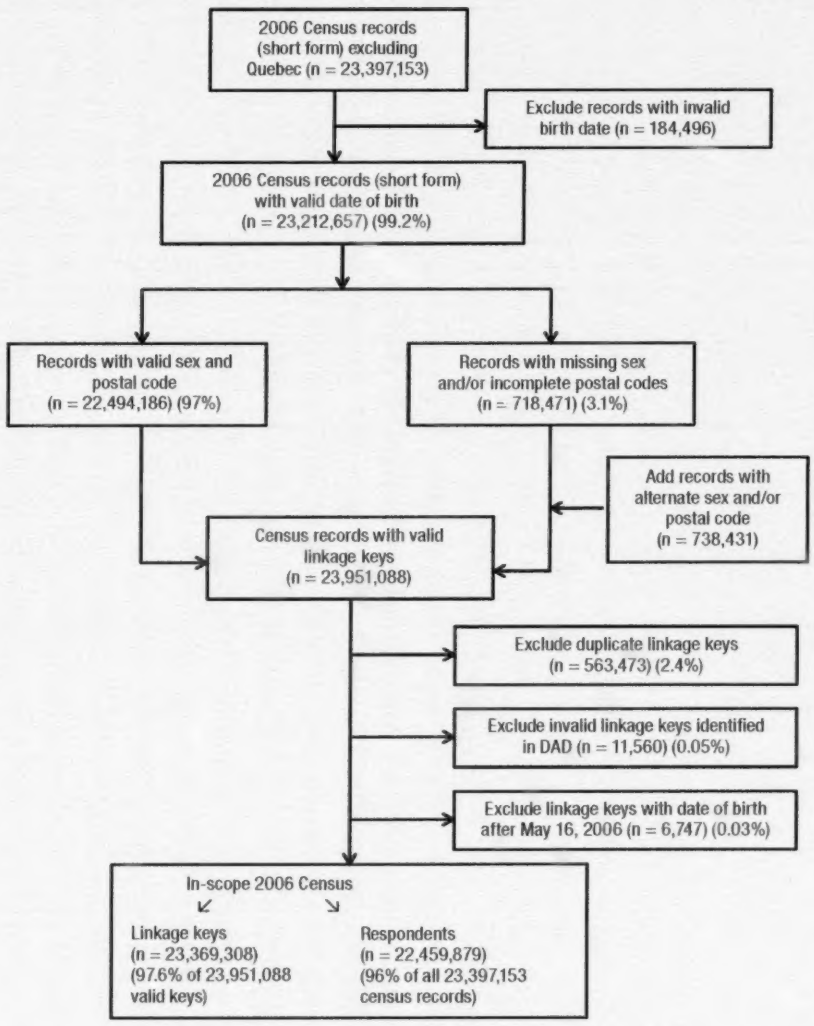
T1 Personal Master Files

The T1 Personal Master File (T1PMF) is an annual file derived from tax returns. It contains name, date of birth, sex and postal code, which can be used for record linkage. To obtain postal code information for census respondents whose postal code was missing, incomplete, or had changed, T1PMFs for 2005 to 2009 were linked deterministically to the 2006 Census (short-form) file, based on sex, birth date, and partial family and first given names; income information was not retained. Approximately 90% of census records linked to at least one T1PMF year. For individuals who did not file taxes annually and/or were not required to file (for example, children), postal codes were assigned based on information about other household members who were tax-filers.

Discharge Abstract Database (DAD)

The DAD contains approximately 3 million hospital records with demographic, administrative and clinical data, as well as HINs, for all acute-care and some psychiatric, chronic, rehabilitation and day-surgery hospital discharges occurring in a fiscal year (April 1 to March 31) in all provinces and territories except Quebec.[26,27] Re-abstraction studies that compare information on the original DAD record with corresponding fields in patient charts repeatedly find that the non-clinical data elements, including date of birth, sex and postal code, are very reliable, with differences observed in these fields amounting to less than 2%.[27]

Figure 1
Processing of the 2006 Census for record linkage


2006 Census records (short form) excluding Quebec (n = 23,397,153)
Exclude records with invalid birth date (n = 184,496)
2006 Census records (short form) with valid date of birth (n = 23,212,657) (99.2%)
Records with valid sex and postal code (n = 22,494,186) (97%)
Records with missing sex and/or incomplete postal codes (n = 718,471) (3.1%)
Add records with alternate sex and/or postal code (n = 738,431)
Census records with valid linkage keys (n = 23,951,088)
Exclude duplicate linkage keys (n = 563,473) (2.4%)
Exclude invalid linkage keys identified in DAD (n = 11,560) (0.05%)
Exclude linkage keys with date of birth after May 16, 2006 (n = 6,747) (0.03%)
In-scope 2006 Census
Linkage keys (n = 23,369,308) (97.6% of 23,951,088 valid keys)
Respondents (n = 22,459,879) (96% of all 23,397,153 census records)

For the purposes of record linkage, DAD data from 2005/2006 to 2008/2009 were used in the pre-processing phase of this study; data from 2006/2007 to 2008/2009 were used in the data linkage phase (2006/2007 n = 3,186,079; 2007/2008 n = 3,204,838; 2008/2009 n = 3,232,396). Because records pertain to hospitalizations, not to people, individuals who were hospitalized several times can be represented more than once in the DAD.

Methods

Record linkage involved three steps: data processing, record linkage and validation.

Data processing

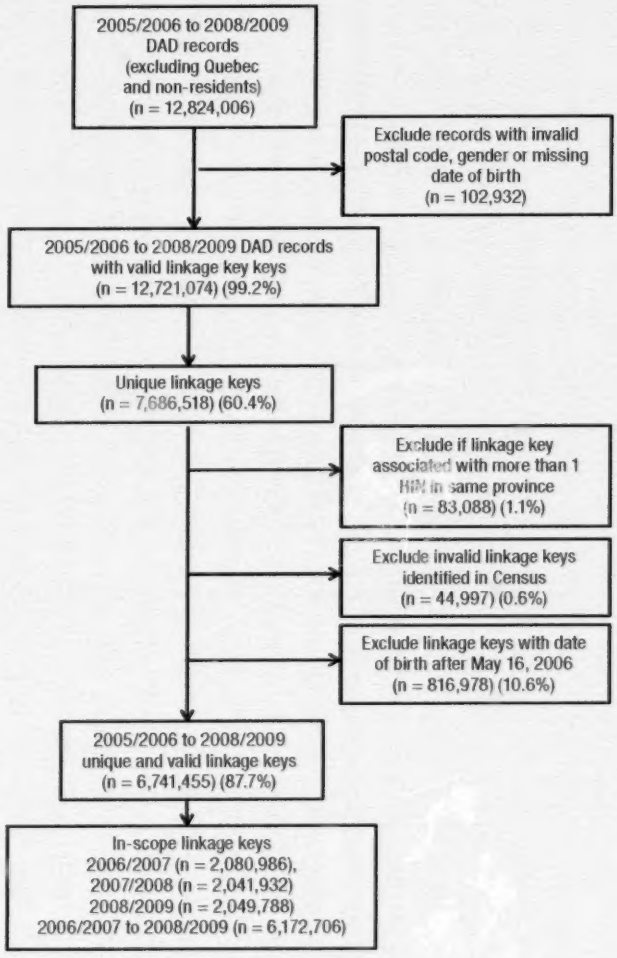
Before linkage, the data were processed to improve the quality of the variables

that would be used (date of birth, postal code and sex) and to establish the unique set of linkage keys in each data file to minimize false links. This processing identified data errors or omissions that could result in false links.

A total of 23,397,153 census records were available for linkage[28] (Figure 1). Quebec records were excluded because Statistics Canada does not have access to the corresponding hospitalization data. Census records with an invalid or incomplete birth date were excluded. Records with missing information for sex were assigned a sex, and a duplicate record was created with the same birth date and postal code, but the opposite sex. Records with differing postal code information in the original and post-processed fields were also duplicated: one record with the original postal code; the other with the post-processed code. Original and duplicate census records were associated using a census group identifier, enabling the identification and removal of the duplicates once the linkage was completed. Lastly, a series of exclusions was applied to the census records to establish the final set of valid and unique linkage keys. The following keys were excluded: duplicate keys with the same date of birth, postal code and sex (for example, same-sex twins living in the same location); invalid linkage keys identified as a result of DAD data processing (described below); and keys with birthdates after May 16, 2006 (Census Day). In total, 23,369,308 valid and unique census keys were in-scope for data linkage, representing 96% of census respondents (excluding Quebec).

DAD records for fiscal years 2005/2006 to 2008/2009 pertaining to Canadian residents hospitalized in provinces and territories outside Quebec were eligible for pre-processing (12,824,006) (Figure 2). Records with invalid or missing birth date, sex or postal code were excluded. The remaining DAD records underwent additional processing to improve consistency and accuracy of HINs and to ensure a one-to-one correspondence with the linkage keys. A common adjustment was the replacement of a temporary (typically the mother's) or missing HIN on an infant's DAD record with a permanent HIN that had subsequently been assigned and that appeared on subsequent hospitalization records.

Figure 2
Processing of the 2005/2006 to 2008/2009 Discharge Abstract Database for record linkage


2005/2006 to 2008/2009 DAD records (excluding Quebec and non-residents) (n = 12,824,006)
Exclude records with invalid postal code, gender or missing date of birth (n = 102,932)
2005/2006 to 2008/2009 DAD records with valid linkage key keys (n = 12,721,074) (99.2%)
Unique linkage keys (n = 7,686,518) (60.4%)
Exclude if linkage key associated with more than 1 HIN in same province (n = 83,088) (1.1%)
Exclude invalid linkage keys identified in Census (n = 44,997) (0.6%)
Exclude linkage keys with date of birth after May 16, 2006 (n = 816,978) (10.6%)
2005/2006 to 2008/2009 unique and valid linkage keys (n = 6,741,455) (87.7%)
In-scope linkage keys
2006/2007 (n = 2,080,986),
2007/2008 (n = 2,041,932)
2008/2009 (n = 2,049,788)
2006/2007 to 2008/2009 (n = 6,172,706)

More than 12.7 million DAD records with valid linking information were available, representing 7,686,518 unique and valid linkage keys (Figure 2). The large number of duplicate linkage keys was expected, given that individuals could be admitted to hospital more than once. The following keys were excluded: linkage keys associated with multiple HINs in the same province; invalid keys identified as a result of census data processing described above; and keys with birthdates after Census Day. In total, 6,172,706 valid and unique linkage keys for fiscal years 2006/2007 to 2008/2009 were in-scope for data linkage.

Record linkage

The hierarchical deterministic exact matching approach to link census and DAD data involved an iterative process in which linkage keys comprised of date of birth, sex and postal code were compared across files. The use of multiple

keys applied consecutively maximizes the discriminating power of the linking information and minimizes the impact of errors and missing values.[13]

The iterative approach applied 28 rules. Early iterations observed stringent rules; subsequent passes tolerated some divergence (Table 1). For example, the first iteration required an exact match between the census birth date, sex and postal code and the DAD linkage keys. Iterations 2 to 4 required an exact match on birth date, sex and postal code obtained from the T1PMF and the DAD linkage keys. Postal codes from the 2005, 2006 and 2007 tax files were used when the census file was linked to the 2006/2007 DAD data. The 2006-to-2008 and 2007-to-2009 tax files were used when the census was linked to the 2007/2008 and 2008/2009 DAD data, respectively. Iterations 5 to 10 relaxed the rules for postal code, allowing one of the six characters of the census-reported postal code to be dropped. The process was repeated using the T1PMF postal codes (iterations 11 to 28).

At each iteration, only unique linkage keys from one file were compared to the unique keys in the other. When the iteration was completed, linked keys were removed from future iterations to ensure that keys were linked only once. Linked keys that had been added to deal with missing sex and/or postal code and sharing the same census group identifier as the linked keys were also removed. For census keys linking to a DAD key, a deterministic linkage to the full set of DAD discharge records was conducted using the linkage key, related HIN, and issuing province. Data processing and record linkage were conducted using SAS 9.2.

Validation

Two types of linkage rates are reported. First, the percentage of DAD keys linking to census keys is reported for each iteration of the linkage. The final linkage rate is not expected to be 100%, owing to differences in the populations represented in the census and DAD, census undercoverage of specific subpopulations, and use of hospital services by individuals who entered Canada after Census Day (for example, new immigrants). Lower linkage rates are also expected among people who were institutionalized (for example, residents of long-term care facilities), since residents would share the same postal code, thereby reducing the uniqueness of the linkage keys.

Second, the percentage of census records eligible for linkage that linked to the DAD (2006/2007 to 2008/2009) is reported by province/territory and by selected socio-demographic characteristics. These rates were based on long-form census respondents (the validation cohort) eligible for record linkage (4,652,683 records; 94% of all long-form respondents) and reflect the prevalence of being hospitalized at least once in the respective fiscal year. Linkage rates were expected to be higher among groups, such as seniors, who are more likely to be hospitalized than non-seniors. Variation in linkage rates across provinces/territories may also reflect differences in health care delivery.[29]

Coverage analysis

When only a fraction of records are expected to link, evaluating the quality of a data linkage is challenging. This is typical of health-based data linkages, where, for example, limited numbers of individuals are expected to be hospitalized or to die during the follow-up period. In such situations, it is not obvious if unlinked records represent missed links or if the event of interest did not occur. The quality of such linkages has been assessed by comparing results of different approaches that used the same data[16,17,30-33] and by comparing outcome rates and percentage distributions of variables available in linked and unlinked data.[30,34,35]

Annual national (excluding Quebec) and jurisdiction-specific unweighted and weighted coverage rates were calculated by dividing the number of acute-care hospital discharges among long-form census respondents in each jurisdiction according to the linked census-DAD data (numerator) by the number of acute-care hospital discharges reported in the unlinked 2006/2007, 2007/2008 and 2008/2009 DAD data (denominator). To more closely match the target population of the linked data, where possible, DAD records pertaining to populations

What is already known on this subject?

- Record linkage is a cost-effective way to add value to administrative and survey datasets.
- Combining complementary data sources permits analyses that would otherwise not be possible.
- The use of provincial health insurance registries to facilitate record linkages for population health research is well established.
- Registry and non-registry approaches to linking health administration and 2006 Census information have yielded research-quality data for the provinces of Ontario and Manitoba for 2006/2007.

What does this study add?

- This study presents the results of a hierarchical deterministic exact matching approach to link 2006 Census data to 2006/2007-to-2008/09 hospitalization records in the Discharge Abstract Database (DAD) for all Canadian jurisdictions (excluding Quebec,) based on date of birth, postal code and sex.
- Use of tax data helped to ensure that the postal codes of census respondents were up-to-date, thereby minimizing missed links.
- The linked cohort represents approximately 80% of hospitalizations that occurred during the 2006/2007-to-2008/2009 period.
- The nationally representative linked census-DAD file fills a data gap and offers new research opportunities.

Table 1
Cumulative percentage and number of Discharge Abstract Database (DAD) keys linked to 2006 Census, by iteration, Canada excluding Quebec, 2006/2007, 2007/2008 and 2008/2009

	DAD year 2006/2007			DAD year 2007/2008			DAD year 2008/2009		
	Source of postal code (census key only) and format	Cumulative		Source of postal code (census key only) and format	Cumulative		Source of postal code (census key only) and format	Cumulative	
Postal code used/Iteration	if incomplete	%	Number	if incomplete	%	Number	if incomplete	%	Number
Full census postal code (FSA LDU) were used									
1	census	71.8	1,495,004	census	65.1	1,329,408	census	59.4	1,217,899
Full T1PMF postal code (FSA LDU) were used									
2	2005 T1PMF	74.3	1,545,888	2006 T1PMF	72.0	1,470,679	2007 T1PMF	70.3	1,440,860
3	2006 T1PMF	75.5	1,570,658	2007 T1PMF	73.8	1,506,694	2008 T1PMF	72.0	1,475,942
4	2007 T1PMF	75.7	1,574,321	2008 T1PMF	74.1	1,512,212	2009 T1PMF	72.3	1,482,647
Partial postal codes from census were used (x indicates alpha-numeric value dropped from key)									
5	FSA LDx	76.5	1,591,750	FSA LDx	74.8	1,527,771	FSA LDx	73.0	1,496,968
6	FSA LxU	77.9	1,621,704	FSA LxU	76.3	1,557,111	FSA LxU	74.4	1,525,707
7	FSA xDU	78.2	1,627,811	FSA xDU	76.5	1,562,431	FSA xDU	74.7	1,530,638
8	FSx LDU	78.9	1,642,783	FSx LDU	77.2	1,575,974	FSx LDU	75.3	1,543,716
9	FxA LDU	79.0	1,644,879	FxA LDU	77.3	1,577,857	FxA LDU	75.4	1,545,438
10	xSA LDU	79.2	1,648,534	xSA LDU	77.5	1,581,599	xSA LDU	75.6	1,549,564
Partial postal codes from T1PMF (2005, 2006, 2007) pertaining to DAD year-1 were used									
11	FSA LDx	79.3	1,649,501	FSA LDx	77.6	1,583,960	FSA LDx	75.8	1,553,001
12	FSA LxU	79.4	1,651,674	FSA LxU	77.8	1,587,924	FSA LxU	76.0	1,558,490
13	FSA xDU	79.4	1,652,018	FSA xDU	77.8	1,588,708	FSA xDU	76.1	1,559,725
14	FSx LDU	79.4	1,652,996	FSx LDU	77.9	1,590,642	FSx LDU	76.2	1,562,475
15	FxA LDU	79.4	1,653,106	FxA LDU	77.9	1,590,959	FxA LDU	76.2	1,562,937
16	xSA LDU	79.5	1,653,459	xSA LDU	77.9	1,591,447	xSA LDU	76.3	1,563,595
Partial postal codes from T1PMF (2006, 2007, 2008) pertaining to DAD year were used									
17	FSA LDx	79.5	1,654,162	FSA LDx	78.0	1,592,392	FSA LDx	76.3	1,564,602
18	FSA LxU	79.6	1,655,504	FSA LxU	78.1	1,593,978	FSA LxU	76.4	1,566,193
19	FSA xDU	79.6	1,655,723	FSA xDU	78.1	1,594,306	FSA xDU	76.4	1,566,493
20	FSx LDU	79.6	1,656,310	FSx LDU	78.1	1,595,077	FSx LDU	76.5	1,567,262
21	FxA LDU	79.6	1,656,404	FxA LDU	78.1	1,595,219	FxA LDU	76.5	1,567,402
22	xSA LDU	79.6	1,656,620	xSA LDU	78.1	1,595,467	xSA LDU	76.5	1,567,654
Partial postal codes from T1PMF (2007, 2008, 2009) pertaining to DAD year +1 were used									
23	FSA LDx	79.6	1,657,049	FSA LDx	78.2	1,595,918	FSA LDx	76.5	1,568,170
24	FSA LxU	79.7	1,657,975	FSA LxU	78.2	1,596,983	FSA LxU	76.6	1,569,308
25	FSA xDU	79.7	1,658,107	FSA xDU	78.2	1,597,128	FSA xDU	76.6	1,569,469
26	FSx LDU	79.7	1,658,563	FSx LDU	78.2	1,597,581	FSx LDU	76.6	1,569,988
27	FxA LDU	79.7	1,658,619	FxA LDU	78.2	1,597,653	FxA LDU	76.6	1,570,075
28	xSA LDU	79.7	1,658,817	xSA LDU	78.3	1,597,868	xSA LDU	76.6	1,570,329

Notes: Postal codes have six digits. The first three—FSA—pertain to the Forward Sortation Area; the last three—LDU—pertain to the Local Delivery Unit. An "x" in a postal code (FSA LDU) indicates the digit dropped from the linkage key. Each key contains birth date, sex and postal code. DAD keys used the postal code of the patient's residence and were constant over the 28 iterations. Census keys used postal codes from the census and alternates from tax files (T1PMF); postal code source and completeness changed with each iteration. The DAD keys eligible for linkage were: 2006/2007 = 2,080,986; 2007/2008 = 2,041,932; and 2008/2009 = 2,049,788.

not captured by the long-form census were removed from the denominator: residents of seniors' homes, people born after Census Day, stillbirths, and non-Canadians.

Unweighted coverage rates should approach the percentage of the population completing the long-form census (about 20% nationally, varying from 16% for Newfoundland and Labrador to 63% to 69% for Nunavut).[22,24] Weighted coverage rates should approach, but not equal, 100%, owing to differences in the populations covered by the linked census-DAD data and the unlinked DAD data. For example, the institutionalized population, who are high users of hospital services,[36,37] are represented in the unlinked DAD data, but not in the linked census-DAD data.

Sociodemographic characteristics

Linkage rates were compared by selected socioeconomic characteristics from the census to determine if rates were higher among people more likely to be hospitalized, such as those in lower-income groups and Aboriginal people.[36,38,39]

Income quintiles were derived at the economic family level or directly for unattached individuals.[23] Total after-tax income from all sources and from all family members/individuals was summed, adjusted for family size, and divided into *quintiles*. To minimize regional income differences, quintiles were estimated separately for each province/territory, and then pooled.

Highest level of education of people aged 18 or older was dichotomized as: at least secondary school graduation, or less than secondary graduation. People younger than 18, most of whom would be too young to have graduated, were excluded.

Information on *Aboriginal status* was derived from the question: "Is this person an Aboriginal person, that is, North American Indian, Métis or Inuit (Eskimo)?" Respondents marked all that applied. Responses were grouped into

Table 2
Percentage and number of Discharge Abstract Database (DAD) keys linked to 2006 Census and percentage distribution of keys linked by iteration, by province, sex and age group, Canada excluding Quebec, 2006/2007, 2007/2008 and 2008/2009

	2006/2007					2007/2008					2008/2009				
			Percentage of keys linked, by iteration					Percentage of keys linked, by iteration					Percentage of keys linked, by iteration		
	%	Number	1	2 to 4	5 to 28	%	Number	1	2 to 4	5 to 28	%	Number	1	2 to 4	5 to 28
Canada excluding Quebec	79.7	1,658,817	90.1	4.8	5.1	78.3	1,597,868	83.2	11.4	5.4	76.6	1,570,329	77.6	16.9	5.6
Province/Territory															
Newfoundland and Labrador	84.5	66,832	91.6	3.9	4.6	84.7	66,880	87.9	7.8	4.3	83.8	66,186	83.8	11.6	4.6
Prince Edward Island	83.0	16,162	90.0	4.5	5.5	82.4	15,553	84.9	9.1	6.0	82.3	15,667	81.4	13.2	5.4
Nova Scotia	84.0	99,818	90.9	4.7	4.4	83.3	99,493	86.0	9.5	4.6	82.1	100,570	81.1	14.0	4.9
New Brunswick	80.8	78,694	91.7	5.2	3.1	79.5	74,791	85.5	11.1	3.4	78.1	72,768	80.7	16.2	3.1
Ontario	80.3	577,266	91.1	4.2	4.7	78.6	545,688	83.7	11.3	5.0	76.5	527,613	77.7	16.9	5.3
Manitoba	81.2	119,511	90.3	4.3	5.5	79.9	114,433	84.9	9.6	5.5	79.0	114,755	79.8	14.4	5.8
Saskatchewan	81.6	121,636	89.7	5.2	5.1	80.6	118,397	83.9	10.9	5.2	79.6	118,056	78.5	15.9	5.6
Alberta	77.2	177,466	87.9	6.3	5.7	75.1	168,781	79.0	14.8	6.2	72.6	162,203	72.2	21.2	6.6
British Columbia	78.0	391,169	89.1	5.0	5.8	76.5	383,933	81.7	12.2	6.1	75.0	382,745	76.0	17.9	6.1
Yukon	73.8	2,817	85.6	9.2	5.1	71.6	2,754	78.5	15.2	6.3	69.7	2,742	72.1	21.5	6.4
Northwest Territories /Nunavut	73.1	7,242	84.1	6.0	10.0	73.4	6,887	79.8	10.0	10.2	73.0	6,672	76.9	12.4	10.7
Sex															
Male	79.0	675,424	90.5	4.2	5.4	78.5	652,339	85.2	9.3	5.5	77.6	646,688	80.8	13.7	5.5
Female	80.2	983,393	89.9	5.2	4.9	78.1	945,529	81.8	12.9	5.3	75.9	923,641	75.3	19.1	5.6
Age group															
Younger than 1	73.3	37,835	91.3	3.8	4.9	69.8	9,237	72.9	19.5	7.6	67.5	7,313	63.1	27.9	9.1
1 to 4	78.7	35,894	87.3	6.0	6.7	76.2	29,564	78.0	15.1	6.9	74.1	25,841	70.6	22.2	7.2
5 to 9	81.9	27,034	89.2	4.7	6.2	78.6	22,638	81.9	11.6	6.5	77.1	20,166	75.5	17.5	7.0
10 to 14	81.6	25,870	89.3	4.4	6.2	78.9	28,664	82.6	10.3	7.0	76.1	32,349	77.7	15.2	7.1
15 to 19	72.8	49,831	85.1	6.5	8.4	68.1	52,379	74.5	15.9	9.6	65.0	54,298	66.5	23.2	10.4
20 to 24	67.6	70,637	79.8	11.2	8.9	62.2	70,224	63.1	26.6	10.3	59.0	70,808	51.2	37.6	11.2
25 to 34	77.9	225,732	86.4	8.1	5.5	73.4	217,485	71.6	22.2	6.2	69.9	205,788	60.0	33.1	6.9
35 to 44	80.1	191,606	89.0	5.2	5.7	77.7	180,773	80.9	12.9	6.2	76.0	173,808	74.5	19.1	6.4
45 to 54	82.5	214,923	90.7	4.0	5.3	81.7	218,754	85.6	9.0	5.4	81.0	224,727	81.2	13.3	5.5
55 to 64	84.6	239,176	92.0	3.6	4.4	84.5	245,326	87.9	7.8	4.3	83.9	252,859	84.1	11.7	4.2
65 to 74	84.0	234,975	93.2	3.1	3.7	84.3	237,871	89.9	6.5	3.6	83.8	240,058	86.8	9.6	3.6
75 or older	77.5	305,304	92.9	3.1	4.0	78.0	284,953	89.5	6.3	4.1	76.4	262,314	86.2	9.7	4.1

Notes: Sex and age group were defined using information available in the DAD and available in the linkage key only (comprised of birth date//sex//postal code). Province/Territory was defined based on the first letter of the postal code. Because their postal codes all begin with X, the Northwest Territories and Nunavut were combined.
Sources: 2006 Census of Population; 2006/2007 to 2008/2009 census-linked Discharge Abstract Database.

six categories: North American Indian (only), Métis (only), Inuit (only), other Aboriginal (multiple or indeterminate), Aboriginal (composite of four preceding categories), or non-Aboriginal.

Country of birth, citizenship, and immigration status were combined into an *immigrant status* variable: immigrant, non-immigrant or non-permanent resident. Immigrants were further subdivided into *long-term* (arrived at least 10 years before the 2006 Census) and *recent* (arrived in the 9 years before the 2006 Census).

A *one-year residential mobility* variable was created to reflect address changes: same address, moved within Canada, or moved from outside Canada. This was derived by comparing each respondent's municipality and province of residence on Census Day and one year earlier.

A *rural/urban* variable reflected location and community size. Farm and non-farm residences in areas with a population of less than 1,000 were considered rural/farm. Population centres were categorized as small (1,000 to 29,999), medium (30,000 to 99,999) and large (100,000 or more).

Respecting respondent privacy

Statistics Canada ensures respondent privacy during the linkage and subsequent use of linked files. Only employees directly involved in the process have access to the unique identifying information required for linkage (such as names and HINs) and do not access health-related information. When the data linkage is completed, an analytical file is created from which identifying information is removed. This de-identified file is accessed by analysts for validation and analysis.

Results

Linkage results

Overall, 80%, or 1.66 million of eligible 2006/2007 DAD keys, were linked to the 2006 Census (Table 1). Results were similar for the eligible 2007/2008 (78% or 1.60 million) and 2008/2009 (77% or 1.57 million) DAD keys. The

Table 3
Number and percentage of long-form 2006 Census respondents eligible for linkage and discharged from acute-care hospitals, by selected characteristics, Canada excluding Quebec, 2006/2007, 2007/2008 and 2008/2009

	Long-form census respondents	Eligible for linkage		Eligible census respondents linked to DAD					
				2006/2007		2007/2008		2008/2009	
	'000	'000	%	'000	%	'000	%	'000	%
Canada excluding Quebec	4,946.2	4,652.7	94	259.7	5.6	249.1	5.4	241.5	5.2
Province/Territory									
Newfoundland and Labrador	103.5	97.8	94	5.9	6.1	5.9	6.0	5.7	5.8
Prince Edward Island	26.2	24.5	94	1.8	7.2	1.7	6.9	1.6	6.6
Nova Scotia	182.4	171.6	94	9.7	5.6	9.4	5.5	9.2	5.4
New Brunswick	146.5	141.2	96	10.0	7.1	9.5	6.8	9.0	6.4
Ontario	2,387.9	2,254.5	94	114.2	5.1	109.3	4.8	105.9	4.7
Manitoba	278.9	263.7	94	17.4	6.6	16.5	6.3	16.6	6.3
Saskatchewan	239.2	225.2	94	17.6	7.8	16.5	7.3	15.9	7.1
Alberta	681.5	640.7	94	37.3	5.8	35.8	5.6	34.4	5.4
British Columbia	830.5	769.1	92	41.6	5.4	40.4	5.3	39.2	5.1
Yukon	13.9	13.0	94	0.7	5.3	0.7	5.5	0.7	5.7
Northwest Territories	26.3	24.6	93	1.7	6.8	1.6	6.5	1.5	6.2
Nunavut	29.3	26.8	91	1.7	6.5	1.7	6.4	1.7	6.4
Sex									
Male	2,432.5	2,284.4	94	101.8	4.5	97.5	4.3	95.5	4.2
Female	2,513.7	2,368.3	94	157.9	6.7	151.5	6.4	146.0	6.2
Age group									
Younger than 1	56.8	51.8	91	8.4	16.2	1.7	3.4	1.1	2.1
1 to 4	229.8	213.0	93	5.9	2.7	4.2	2.0	3.4	1.6
5 to 9	305.2	284.4	93	4.2	1.5	3.7	1.3	3.4	1.2
10 to 14	345.1	321.7	93	4.7	1.5	5.5	1.7	6.1	1.9
15 to 19	354.7	330.7	93	9.8	2.9	10.4	3.2	10.9	3.3
20 to 24	329.6	307.7	93	13.8	4.5	14.1	4.6	13.7	4.5
25 to 34	623.4	587.4	94	41.7	7.1	40.2	6.8	37.2	6.3
35 to 44	756.2	715.5	95	29.1	4.1	26.7	3.7	24.9	3.5
45 to 54	767.2	725.0	94	28.8	4.0	28.9	4.0	29.3	4.0
55 to 64	556.1	525.3	94	33.0	6.3	34.0	6.5	34.1	6.5
65 to 74	343.6	325.9	95	34.5	10.6	35.2	10.8	35.3	10.8
75 or older	278.5	264.2	95	46.0	17.4	44.4	16.8	42.0	15.9
Household income quintile									
1 (lowest)	989.2	911.6	92	56.7	6.2	53.0	5.8	50.6	5.6
2	989.2	927.1	94	58.7	6.3	56.5	6.1	54.6	5.9
3	989.3	933.6	94	51.5	5.5	49.8	5.3	48.5	5.2
4	989.3	938.1	95	47.8	5.1	46.2	4.9	44.9	4.8
5 (highest)	989.3	942.1	95	44.9	4.8	43.6	4.6	42.9	4.6
Education (aged 18 or older)									
Secondary graduation or more	2,979.1	2,818.5	95	160.0	5.7	158.1	5.6	153.6	5.5
Less than secondary graduation	811.9	759.9	93	71.2	9.4	69.8	9.2	67.5	8.9
Aboriginal identity									
Any Aboriginal	470.2	422.2	90	32.6	7.7	31.0	7.3	30.2	7.2
North American Indian only	345.0	306.4	89	25.0	8.1	23.8	7.8	23.3	7.6
Métis only	84.3	78.4	93	5.0	6.4	4.6	5.9	4.5	5.7
Inuit only	32.5	29.9	92	2.1	7.0	2.0	6.9	2.0	6.7
Mulitple Aboriginal	8.4	7.6	91	0.5	6.8	0.5	6.9	0.5	6.6
Non-Aboriginal	4,475.9	4,230.4	94	227.1	5.4	218.1	5.2	211.3	5.0
Immigrant status									
Non-immigrant	3,864.2	3,627.2	94	205.5	5.7	196.2	5.4	190.3	5.2
Immigrant	1,034.8	982.5	95	53.1	5.4	51.9	5.3	50.2	5.1
Long-term (arrived 10 or more years before 2006 Census)	315.5	297.9	94	11.9	4.0	11.1	3.7	10.3	3.4
Recent (arrived 9 or fewer years before 2006 Census)	719.3	684.6	95	41.1	6.0	40.8	6.0	40.0	5.8
Non-permanent resident	47.0	42.9	91	1.1	2.5	1.0	2.4	1.0	2.3

majority of links between the DAD and the 2006 Census (72% for 2006/2007 DAD to 59% for 2008/2009 DAD, or 1.50 to 1.22 million, respectively) were achieved in the first iteration, which required an exact match on birth date, sex and postal code. The number of links achieved using postal codes from tax files (iterations 2 to 4) ranged from 79,000 (4%) with the 2006/2007 DAD to 265,000 (13%) with the 2007/2008 DAD. Iterations 5 to 28 added an additional 85,000 to 88,000 (4%) links, depending on the DAD year.

The percentage of DAD keys linking to census keys tended to be consistent across provinces and age groups, but some exceptions were evident (Table 2). For example, the percentages of DAD keys that linked were comparatively low in Alberta (77% with 2006/2007 DAD to 73% with 2008/2009), British Columbia (78% to 75%) and the territories (73% to 72%). Lower rates were also observed for infants younger than age 1 (73% to 68%) and for 15- to 24-year-olds (70% to 61%).

The percentage of long-form census respondents who linked to the DAD (that is, they were hospitalized) ranged from 5.6% (2006/2007) to 5.2% (2008/2009) (Table 3). Linkage rates reflected expected differential use of hospital services. The rate was higher among females than males. Infants younger than 1 on Census Day (May 16, 2006) and seniors were more likely than other age groups to link to 2006/2007 DAD records. In subsequent DAD years, seniors' linkage rate remained comparatively high, but the rate among children who had been younger than age 1 on Census Day fell to that of children who had been aged 1 to 4 on Census Day. Other groups with higher linkage rates were people in low-income quintiles (6%) and those who identified as Aboriginal (7%). Linkage rates tended to be higher among rural than among urban populations.

Table 3 (concluded)
Number and percentage of long-form 2006 Census respondents eligible for linkage and discharged from acute-care hospitals, by selected characteristics, Canada excluding Quebec, 2006/2007, 2007/2008 and 2008/2009

	Long-form census respondents	Eligible for linkage		Eligible census respondents linked to DAD					
				2006/2007		2007/2008		2008/2009	
	'000	'000	%	'000	%	'000	%	'000	%
Residential mobility									
Same address	4,181.3	3,943.4	94	217.5	5.5	214.5	5.4	208.6	5.3
Different Canadian address	651.9	610.5	94	32.2	5.3	31.3	5.1	30.3	5.0
Moved from outside Canada	50.5	46.8	93	1.6	3.5	1.6	3.3	1.4	3.1
Residence									
Rural and farm	1,213.0	1,112.3	92	75.3	6.8	73.1	6.6	70.7	6.4
Population centre									
Small (1,000 to 29,999)	600.3	560.6	93	38.2	6.8	36.8	6.6	35.5	6.3
Medium (30,000 to 99,999)	425.6	405.4	95	24.7	6.1	23.5	5.8	22.9	5.7
Large (100,000 or more)	2,707.2	2,574.5	95	121.4	4.7	115.6	4.5	112.3	4.4

% eligible for linkage = (number of long-form census respondents with unique linkage key/number of long-form census respondents)*100
% hospitalized = (number of eligible long-form census respondents hospitalized/number of eligible long-form census respondents)*100
Sources: 2006 Census of Population; 2006/2007 to 2008/2009 census-linked Discharge Abstract Database (DAD).

Coverage evaluation

Coverage rates for all-cause hospitalizations for 2006/2007 to 2008/2009 were 17% (unweighted) and 80% to 78% (weighted) (Table 4), but varied by jurisdiction. For example, unweighted coverage rates ranged from 16% to 22% in the provinces, and from 29% to 69% in the territories. Weighted coverage rates ranged from 75% to 84% in the provinces, and from 62% to 72% in the territories.

Over the study period, unweighted and weighted coverage rates were similar by sex, but not by age group. The weighted 2006/2007 rates for infants younger than age 1 and 15- to 24-year-olds were 6 to 10 percentage points below the all-ages total. Weighted coverage rates based on the 2007/2008 and 2008/2009 linked files also reflected this pattern, but because age was defined by the census rather than by hospitalization, undercoverage of youth in the later DAD files was apparent in the next-oldest age groups.

Discussion

Based on a hierarchical deterministic exact matching approach, about 80% of linkage keys identified in the hospitalization data were linked to the 2006 Census. This was similar to other Canadian studies, which reported match rates of 75% among records expected to link.[5,40,41]

The hierarchical approach identified matches that would have been missed by a deterministic exact match approach conducted in a single pass.[13] While the majority of links in the present study were made in the first iteration, an additional 8% to 17% were made in subsequent iterations. The use of updated postal codes from tax data to account for mobility contributed to achieving additional links, particularly for later years, and overcame a limitation typical of most census-follow-up studies.[42] Linkage rates, coverage and quality of the data linkage remained consistent throughout the three-year study period.

Linkage rates to the DAD among the census long-form cohort were 5% to 6%, representing the percentage who experienced at least one hospitalization during the three years. Linkage rates were higher among specific groups: seniors, people in low income quintiles, and Aboriginal people. This is consistent with previous research,[36,38,39] and provides evidence of the validity of the linkage and the suitability of the linked data for health analysis.

Coverage analysis revealed that the linked census-DAD file represents the majority of hospital events that occurred during the period (weighted results). Furthermore, the pattern of hospitalizations by patient characteristics was similar to results from the DAD alone, with some exceptions.

Unweighted coverage rates varied geographically, with higher rates in the territories, Manitoba's and Saskatchewan. This is attributable to the sampling strategy of the census, whereby up to 100% of individuals in remote areas and on Indian Reserves and Settlements were asked to complete the long-form questionnaire. When weights were applied, Manitoba's and Saskatchewan's coverage rates were closer to rates for the other provinces. However, weighted rates for the territories were lower than those for the provinces; linked data may underestimate hospitalizations of territorial residents because of census undercoverage, higher rates of mobility, and/or a tendency to be hospitalized outside their jurisdiction of residence.[22,24]

As expected, the analysis revealed potential undercoverage of hospitalizations of specific age groups. For infants younger than age 1, lower linkage and coverage rates may be related to the availability of HINs on hospital birth records.[43] Despite attempts to correct this during data processing, some likely remained unresolved. Coverage rates were also lower among 15- to 24-year-olds, possibly reflecting census undercoverage of populations with less stable living arrangements and/or incomplete coverage of some Aboriginal populations.[22,41] Because some of these populations have relatively high hospitalization rates, missed links may have a greater impact on coverage than the absolute numbers would suggest.[38]

Same-sex twins residing at the same address on Census Day and same-sex twins hospitalized with the same postal code would have been dropped from the linked data file because of their non-unique linkage keys. However, twins and higher-order births represent about 3% of all births annually, and same-sex higher-order births are even rarer.[45]

Table 4
Unweighted and weighted coverage rates for acute-care hospital discharges, by province, sex and age, Canada excluding Quebec, 2006/2007, 2007/2008 and 2008/2009

	2006/2007					2007/2008					2008/2009				
	Unlinked hospital discharges (denominator)	Long-form census repondents hospitalized (numerator)				Unlinked hospital discharges (denominator)	Long-form census repondents hospitalized (numerator)				Unlinked hospital discharges (denominator)	Long-form census repondents hospitalized (numerator)			
		Unweighted		Weighted			Unweighted		Weighted			Unweighted		Weighted	
	'000	'000	%	'000	%	'000	'000	%	'000	%	'000	'000	%	'000	%
Canada excluding Quebec	2,045.0	351.8	17.2	1,629.4	79.7	1,967.5	338.2	17.2	1,566.0	79.6	1,942.1	327.0	16.8	1,513.4	77.9
Province/Territory															
Newfoundland and Labrador	52.9	8.4	16.0	40.5	76.5	51.0	8.4	16.4	39.8	78.0	49.4	8.0	16.3	38.3	77.6
Prince Edward Island	15.8	2.5	16.0	12.9	81.6	15.1	2.5	16.5	12.6	83.7	14.8	2.3	15.5	11.7	79.6
Nova Scotia	80.4	13.3	16.5	65.8	81.8	76.7	12.7	16.6	63.1	82.3	77.0	12.5	16.2	62.4	81.0
New Brunswick	87.1	14.2	16.3	69.7	80.0	82.2	13.4	16.3	66.1	80.4	79.3	12.6	15.9	61.8	78.0
Ontario	917.7	149.5	16.3	741.8	80.8	876.3	143.0	16.3	711.3	81.2	863.7	138.6	16.0	688.0	79.7
Manitoba	113.8	25.3	22.2	91.6	80.4	109.6	24.0	21.9	86.6	79.0	109.0	23.6	21.7	84.9	77.9
Saskatchewan	121.6	25.9	21.3	98.8	81.3	115.7	24.5	21.2	92.7	80.1	113.3	23.6	20.8	89.5	79.0
Alberta	291.0	49.7	17.1	227.4	78.2	282.9	48.0	17.0	219.4	77.6	279.3	46.0	16.5	210.9	75.5
British Columbia	352.6	57.4	16.3	273.1	77.4	346.5	56.1	16.2	266.6	76.9	345.0	54.4	15.8	257.8	74.7
Yukon	3.3	1.0	29.0	2.0	62.4	3.3	1.0	31.4	2.2	67.7	3.4	1.0	30.6	2.4	71.5
Northwest Territories	5.1	2.3	45.8	3.3	65.7	5.1	2.3	45.7	3.3	65.8	4.7	2.1	45.0	3.3	69.5
Nunavut	3.7	2.3	63.0	2.4	65.7	3.2	2.2	67.4	2.2	69.4	3.2	2.2	69.0	2.3	71.6
Sex															
Male	858.5	145.6	17.0	679.1	79.1	816.5	139.9	17.1	652.6	79.9	804.4	136.6	17.0	638.0	79.3
Female	1,186.5	206.2	17.4	950.3	80.1	1,151.0	198.3	17.2	913.5	79.4	1,137.7	190.4	16.7	875.4	76.9
Age group															
Younger than 1	54.3	8.9	16.5	39.1	72.0	0.2	0.0	12.6	0.1	55.4					
1 to 4	43.9	8.1	18.4	34.2	78.0	34.7	5.9	17.1	25.4	73.2	20.4	3.3	16.2	14.3	70.3
5 to 9	27.4	5.1	18.7	22.7	82.9	26.4	5.0	18.8	21.9	83.0	25.3	4.5	17.7	19.5	77.3
10 to 14	29.7	5.6	18.7	23.9	80.3	29.3	5.5	18.8	23.4	80.0	28.7	5.1	17.7	21.8	75.9
15 to 19	63.7	11.9	18.7	47.4	74.3	63.9	12.0	18.7	47.1	73.6	62.5	11.7	18.7	45.2	72.3
20 to 24	99.4	16.8	16.9	69.8	70.2	99.7	15.7	15.7	63.0	63.2	97.1	14.6	15.0	57.4	59.1
25 to 34	284.1	48.7	17.2	223.3	78.6	288.8	46.8	16.2	214.4	74.2	286.5	43.4	15.2	196.6	68.6
35 to 44	211.8	36.6	17.3	168.7	79.7	207.8	34.9	16.8	159.7	76.9	202.1	32.8	16.3	149.1	73.8
45 to 54	222.1	38.9	17.5	178.8	80.5	222.3	38.9	17.5	179.0	80.5	221.7	38.0	17.1	175.7	79.2
55 to 64	261.5	47.4	18.1	220.2	84.2	264.0	48.1	18.2	222.5	84.3	266.5	47.2	17.7	221.0	82.9
65 to 74	289.0	51.6	17.8	245.2	84.9	285.7	51.9	18.2	246.7	86.3	286.4	51.7	18.0	245.3	85.7
75 or older	458.2	72.1	15.7	356.2	77.7	444.7	73.6	16.5	362.8	81.6	445.1	74.7	16.8	367.5	82.6

Unweighted coverage rate = (Total hospitalizations of long-form census respondents/Total hospitalizations according to DAD)*100
Weighted coverage rate = (Total hospitalizations of long-form census respondents with census weights/Total hospitalizations according to DAD)*100
Notes: Unlinked and linked DAD estimates exclude stillbirths, non-acute care, non-Canadians, birth dates after Census day (May 16, 2006), health cards not issued by specific province or territory, including unknown, CA, 99, and patients transfered from seniors' residences.
Province for unlinked DAD uses health_card_prov_code.
Sources: 2006 Census of Population; 2006/2007 to 2008/2009 census-linked Discharge Abstract Database (DAD).

Limitations

The linked census-DAD file has a number of important limitations.

First, to derive a set of unique linkage keys, removal of specific keys from both the census and the DAD was necessary. In census short forms, 97.6% of keys were retained; in census long forms, about 94% of keys were retained. The lower eligibility rate of long-form respondents was most evident among those with lower socioeconomic status, people identifying as Aboriginal, rural/farm residents, and residents of Nunavut and British Columbia. Factors such as inaccurate or incomplete recording of dates of birth and/or other demographic-details contribute to this situation.[5,15,34,46]

Second, the coverage analysis compares hospitalizations identified in the census-DAD linked data with those in the DAD data alone. However, the underlying populations differ. While attempts were made to remove DAD records for people not represented in the census (long-form), some may have been missed.

Finally, Quebec was excluded from this study because hospital data for this province are not available to Statistics Canada. In addition, hospitalizations of non-Quebec residents that occurred in Quebec would not be captured in the linked data.

Conclusion

The nationally representative sample and the statistical power provided by the size and population coverage of the linked census-DAD file offers new opportunities for research. Analysis of annual linkage and coverage rates suggests that the file's completeness and quality remained consistent over time. Investigators who use these linked data should consider the potential impact of the linkage methodology, differences in linkage eligibility, linkage and coverage rates, and population exclusions. ■

References

1. Winkler WE. Record linkage. *Sample Surveys: Design, Methods and Applications* 2009; 29A: 351-80.
2. Lix L, Yogendran M, Burchill M, et al. *Defining and Validating Chronic Diseases: An Administrative Data Approach*. Winnipeg, Manitoba: Manitoba Centre for Health Policy, 2006.
3. Iron K, Zagorski BM, Sykora K, et al. Living and dying in Ontario: An opportunity for improved health information. *ICES Investigative Report*. Toronto: Institute for Clinical Evaluative Sciences, 2008.
4. Leong VW, Morgan S, Wong ST, et al. Registration for public drug benefits across areas of differing ethnic composition in British Columbia, Canada. *BMC Health Services Research* (electronic) 2010; 10(171). Available at: www.biomedcentral.com/1472-6963/10/171
5. Wilkins R, Tjepkema M, Mustard C, et al. The Canadian census mortality follow-up study, 1991 through 2001. *Health Reports* 2008; 19(3): 25-43.
6. Wilkins K, Shields M, Rotermann M. Smokers' use of acute-care hospitals—A prospective study. *Health Reports* 2009; 20(4): 75-84.
7. Holman CD, Bass AJ, Rosman DL, et al. A decade of data linkage in Western Australia: Strategic design, applications and benefits of the WA data linkage system. *Australian Health Review* 2008; 32(4): 766-7.
8. Ng E. The healthy immigrant effect and mortality rates. *Health Reports* 2011; 22(4): 24-30.
9. Hwang SW, Wilkins R, Tjepkema M, et al. Mortality among residents of shelters, rooming houses, and hotels in Canada: 11-year follow-up study. *British Medical Journal* 2009; 339: b4036. doi:101136/bmj.b4036
10. Tjepkema M, Wilkins R. Remaining life expectancy at age 25 and probability of survival to age 75, by socio-economic status and Aboriginal ancestry. *Health Reports* 2011; 22(4): 31-6.
11. Manuel DG, Perez R, Bennett C, et al. *Seven More Years: The Impact of Smoking, Alcohol, Diet, Physical Activity, and Stress on Health and Life Expectancy in Ontario*. An ICES/PHO Report. Toronto: Clinical Evaluative Sciences and Public Health Ontario, 2012.
12. Fransoo R, Martens P, Prior H, et al. *Adult Obesity in Manitoba: Prevalence, Associations, and Outcomes*. Winnipeg, Manitoba: Manitoba Centre for Health Policy, 2011.
13. Bernier J, Nobrega K. Overview of record linkage. *SYMPOSIUM 99: Combining Data from Different Sources, May 1999*. Ottawa: Minister of Industry, 2000.
14. Karmel R, Anderson P, Gibson D, et al. Empirical aspects of records linkage across multiple data sets using statistical linkage keys: the experience of PIAC cohort study. *BMC Health Services Research* 2010; 10: 41. Available at: www.biomedcentral.com/1472-6963/10/41
15. NCSIMG. *Statistical Data Linkage in Community Services Data Collection*. Canberra, Australia: Australian Institute of Health and Welfare, 2004. Available at: http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=10737419651&libID=10737419650
16. Li B, Quan H, Fong A, et al. Assessing record linkage between health care and Vital Statistics databases using deterministic methods. *BMC Health Services Research* (electronic) 2006; 6(48). Available at: www.biomedcentral.com/1472-6963/6/48
17. Australian Institute of Health and Welfare. *Comparing an SLK-based and a Name-based Data Linkage Strategy: An Investigation into the PIAC Linkage*. Data Linkage Series No. 11 (Catalogue CSI 11) Canberra, Australia: Australian Institute of Health and Welfare, 2011.
18. HE Development Team. Methodology for creation of the HES patient ID (HESID). Available at: www.hscic.gov.uk/media/1370/HES-Hospital-Episode-Statistics-Replacement-of-the-HES-patient-ID/pdf/HESID_Methodology.pdf
19. Rotermann M, Sanmartin G, Trudeau R, et al. Two approaches to linking census and hospital data in Canada. *Health Reports* 2014; 25(10): 3-14.
20. Statistics Canada. *Approved Record Linkages*. Available at: www.statcan.gc.ca/record-enregistrement/summ-somm-eng.htm
21. Statistics Canada. *Policy on Record Linkage*. Available at: www.statcan.gc.ca/record-enre-gistrement/policy4-1-politique4-1-eng.htm

22. Statistics Canada. *2006 Census Technical Report: Coverage*. Ottawa: Minister of Industry, 2010. Available at: www12.statcan.gc.ca/census-recensement/2006/ref/rp-guides/rp/coverage-couverture/cov-couv_index-eng.cfm

23. Statistics Canada. *2006 Census of Population*. Available at: www12.statcan.gc.ca/census-recensement/2006/index-eng.cfm

24. Statistics Canada. *2006 Census Technical Report on Sampling and Weighting*. Ottawa: Minister of Industry, 2009. Available at: www12.statcan.gc.ca/census-recensement/2006/ref/rp-guides/rp/sw-ep/pdf/92-568-X2006001-eng.pdf

25. Roberts G. Analyzing census microdata in an RDC: What weight to use? *The Research Data Centres Information and Technical Bulletin* (Statistics Canada, Catalogue 12-002-X) Fall 2012; 5(1). Available at: www.statcan.gc.ca/pub/12-002-x/2012001/article/11642-eng.htm

26. Canadian Institute for Health Information. *Discharge Abstract Database (DAD) Metadata*. Available at: http://secure.cihi.ca/cihi-ext-portal/internet/en/document/types+of+care/hospital+care/acute+care/dad_metadata

27. Canadian Institute for Health Information. *CIHI Data Quality of the 2006-2007, 2007-2008, and 2008-2009 Discharge Abstract Database*. Available at: https://secure.cihi.ca/estore/productSeries.htm?pc=PCC228

28. St-Pierre M, Gagné W. *Création d'un prototype de la base de la population canadienne*. Statistics Canada internal working documentation. Ottawa: Statistics Canada, 2012.

29. Canadian Institute for Health Information. *Health Indicators 2013*. Ottawa: Canadian Institute for Health Information, 2013. Available at: https://secure.cihi.ca/free_products/HI2013_Jan30_EN.pdf

30. Hoving JL, Del Monaco A, MacFarlance E. Methodological issues in linking study participants to Australian cancer registries using different methods: lessons from a cohort study. *Australian and New Zealand Journal of Public Health* 2005; 29(4): 378-82.

31. Australian Bureau of Statistics. *Assessing the Quality of Linking School Enrolment Records to 2011 Census Data: Deterministic Linkage Methods*. Research Paper 1351.0.55.045. Available at: http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/FAD1F2FC23A36CE4CA257C3F00180F64/$File/1351055045_dec%202013.pdf

32. Bass J, Garfield C. Statistical linkage keys: How effective are they? *Proceedings of the Symposium on Health Data Linkage, 2003*. Available at: www.publichealth.gov.au/pdf/reports_papers/symposium_procdngs_2003/bass.pdf

33. Beauchamp A, Tonkin AM, Sundararajan V, et al. Validation of de-identified record linkage to ascertain hospital admissions in a cohort study. *BMC Medical Research Methodology* 2011; 11(42). doi:10.1186/1471-2288-11-42

34. Jebamani LS, Burchill CA, Martens PJ. Using data linkage to identify First Nations Manitobans. *Canadian Journal of Public Health* 2005; 96(1): S28-32.

35. Rotermann M. Evaluation of the coverage of linked Canadian Community Health Survey and hospital inpatient records. *Health Reports* 2009; 20(1): 45-51.

36. Canadian Institute for Health Information. *Health Care in Canada*. Ottawa: Canadian Institute for Health Information, 2011. Available at: http://secure.cihi.ca/freeproducts/HCIC_2011_seniors_report_en.pdf

37. Trottier H, Martel L, Houle C, et al. Living at home or in an institution: What makes the difference for seniors. *Health Reports* 2000; 11(4): 49-61.

38. Carrière G, Garner R, Sanmartin C, et al. Acute-care hospitalizations and Aboriginal identity in Canada, 2001/2002. *Health Analysis Division Working Paper Series* (Catalogue 82-622-X) Ottawa: Statistics Canada, 2010. Available at: www.statcan.gc.ca/pub/82-622-x/82-622-x2010005-eng.pdf

39. Sanmartin C, Khan S, LHAD Research Team. Hospitalizations for ambulatory care sensitive conditions (ACSC): The factors that matter. *Health Analysis Division Working Paper Series* (Catalogue 82-622-X) Ottawa: Statistics Canada, 2011. Available at: www.statcan.gc.ca/pub/82-622-x/82-622-x2011007-eng.pdf

40. Brennan J, Lessard M. *A Record Linkage Example: Linkage of the Citizenship and Immigration (CIC) Landing File to the T1FF*. Statistics Canada Methodology Seminar, April 9, 2013.

41. Houle C, Berthelot JM, David P, et al. Matching census database and Manitoba health care files. *Federal Committee on Statistical Methodology Papers*. Ottawa: Federal Committee on Statistical Methodology, 1997. Available at: http://www.statcan.gc.ca/pub/11f0019m/11f0019m1996091-eng.pdf

42. Blakely T, Woodward A, Salmond C. Anonymous linkage of New Zealand mortality and census data. *Australian and New Zealand Journal of Public Health* 2000; 24(1): 92-5.

43. Canadian Institute for Health Information. Special tabulation: Availability of newborn health card numbers 2006/07 to 2012/13. Received April 16, 2014.

44. Statistics Canada. *Incompletely Enumerated Indian Reserves and Indian Settlements*. Available at: www12.statcan.gc.ca/census-recensement/2006/ref/notes/aboriginal-autochtones-eng.cfm

45. Statistics Canada. *Births 2009*. Available at: www.statcan.gc.ca/pub/84f0210x/84f0210x2009000-eng.pdf

46. Australian Institute of Health and Welfare, Australian Bureau of Statistics. *National Best Practice Guidelines for Data Linkage Activities Relating to Aboriginal and Torres Strait Islander People*. Canberra, Australia: Australian Institute of Health and Welfare, 2012. Available at: www.aihw.gov.au/publication-detail/?id=60129542148